

Time: 03 Hours

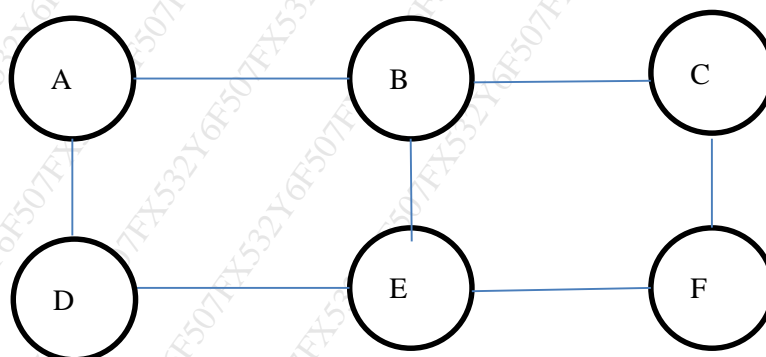
Marks: 80

Note: 1. Question 1 is compulsory

2. Answer any three out of the remaining five questions.

3. Assume any suitable data wherever required and justify the same.

- Q1 a) Explain how big data problems are handled by Hadoop system. [5]
 b) Mention four characteristics of big data and explain in detail. [5]
 c) List and explain the core business drivers behind the NoSQL movement. [5]
 d) Explain the concept of bloom filter with an example. [5]
- Q2 a) What is graph store? Give an example where a graph store can be used to effectively solve a particular business problem. [10]
 b) Write a map reduce pseudo code for word count problem. Illustrate with an example showing all the steps. [10]
- Q3 a) Suppose the stream is $S = \{4, 2, 5, 9, 1, 6, 3, 7\}$. Let hash functions $h(x) = 3x + 7 \pmod{32}$ for some a and b , treat result as a 5-bit binary integer. Show how the Flajolet- Martin algorithm will estimate the number of distinct elements in this stream. [10]
 b) Describe applications of data visualization. [10]
- Q4 a) Explain selection and projection relational algebraic operation using MapReduce. [10]
 b) Explain DGIM algorithm for counting ones in a stream with example. [10]
- Q5 a) Determine communities for the given social network graph using Girvan- Newman algorithm. [10]



b) Consider the following data frame given below: [10]

course	id	class	marks
1	11	1	56
2	12	2	75
3	13	1	48
4	14	2	69
5	15	1	84
6	16	2	53

- i. Create a subset of course less than 5 by using [] brackets and demonstrate the output.
- ii. Create a subset where the course column is less than 4 or the class equals to 1 by using subset () function and demonstrate the output.

Q6 a) i. Write a script to create a dataset named data1 in R containing the following text. [10]

Text: 2, 3, 4, 5, 6.7, 7, 8.1, 9

- ii. Explain the various functions provided by R to combine different sets of data.

b) Describe collaborative filtering in recommendation system. [10]
